

The effects of presenting multidigit mathematics problems in a realistic context on sixth graders' problem solving

Journal:	<i>Cognition and Instruction</i>
Manuscript ID:	HCGI-2012-0048
Manuscript Type:	Original Article
Keywords:	Assessment and Evaluation, Elementary math < Math Education, Quantitative Methodology, Cognitive Modeling

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

**The effects of presenting multidigit mathematics problems in a realistic
context on sixth graders' problem solving**

Date of submission 19-7-2012

*SOLVING MULTIDIGIT MATHEMATICS PROBLEMS***The effects of presenting multidigit mathematics problems in a realistic context on sixth graders' problem solving**

Mathematics education and assessments increasingly involve arithmetic problems presented in context: a realistic situation that requires mathematical modeling. The present study's aim was to assess the effects of such typical school mathematics contexts on two aspects of problem solving, performance and strategy use, with a special focus on the role of students' language level. 685 sixth graders from the Netherlands solved a set of multidigit arithmetic problems on addition, subtraction, multiplication, and division. These problems were presented in two conditions: with and without a realistic context. Regarding performance, item response theory (IRT) models showed first that the same (latent) ability dimension was involved in solving both types of problems. Second, the presence of a context increased the difficulty level of the division problems, but not of the other operations. Regarding strategy use, results showed that strategy choice and strategy accuracy were not affected by the presence of a problem context. More importantly, the absence of context effects on performance and on strategy use held for different subgroups of students, with respect to gender, home language, and language achievement scores. In sum, the present findings suggest that at the end of primary school the presence of a typical context in a multidigit mathematics problem had no marked effects on students' multidigit arithmetic problem solving behavior.

Keywords: word problems, mathematics education, solution strategy, multidimensional IRT, linear logistic test model (LLTM).

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Introduction

Mathematics education has experienced a large international reform (e.g., Kilpatrick, Swafford, & Findell, 2001). A general characteristic of this reform is that instruction no longer focuses predominantly on decontextualized traditional mathematics skills, but that instead the process of solving realistic mathematics problems and doing mathematics are important educational goals (e.g., National Council of Teachers of Mathematics, 1989, 2000). Word problems or the broader category of contextual problems¹ – typically a mathematics structure in a realistic problem situation – serve a central role for several reasons (e.g., Verschaffel, Greer, & De Corte, 2000): they may have motivational potential, mathematical concepts and skills may be developed in a meaningful way, and children may develop knowledge of when and how to use mathematics in everyday-life situations. Furthermore, solving problems in context may ideally serve as a tool for mathematical modeling or mathematizing (e.g., Greer, 1997). As a consequence of this shift in educational goals, mathematics assessments include more and more contextual problems in their tests. For example, the PISA-2009 study (*Programme for International Student Assessment*; OECD, 2010) of students' mathematics performance included mainly problems presented in a real-world situation.

In the Netherlands, the reform is characterized by the principles of Realistic Mathematics Education (RME; Freudenthal, 1973, 1991; Treffers, 1993). In RME, contextual problems (defined as a problem that is experientially real to students) are central: they are the starting point for instruction, which is based on the principle of progressive schematization or mathematization by guided reinvention (Gravemeijer & Doorman, 1999). That is, contextual problems are expected to elicit informal or naive solution strategies which are progressively abbreviated and schematized, in a process guided by the teacher. In the last decades, RME has become the dominant instructional approach in mathematics curricula for Dutch primary education. The most recent national assessments in grade 3 and 6 showed that almost all elementary schools used a mathematics textbook based on RME principles (J. Janssen, Van der Schoot, & Hemker, 2005; Hop, 2012), although a return to more traditionally oriented mathematics textbooks has been observed

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

recently (Royal Netherlands Academy of Arts and Sciences, 2009). These RME-based textbooks contain many contextual problems, although there are substantial differences in this respect between the different textbooks. In order to accommodate these developments, Dutch mathematics assessments (J. Janssen et al.; Hop) and commonly used student monitoring tests also contain predominantly contextual problems. Therefore, today's Dutch primary school students' mathematics education and assessment consists for a large part of problems in realistic contexts.

The growing importance of contextual problems in mathematics education and assessments necessitates that we increase our understanding of the impact of typical school mathematics contexts, both on a theoretical level (what aspects of mathematical cognition are involved?) as well as from a practical educational perspective (what are the implications regarding testing and instruction practices?). Therefore, the main question asked in the current study is: What is the effect on problem solving of presenting a multidigit arithmetic problem in a realistic context typical for school mathematics for students in sixth grade? Two aspects of problem solving are addressed: *performance* (i.e., accuracy) and *solution strategy use*. Furthermore, because a necessary condition for solving the mathematics problem is that the student understands the usually verbal context, it is likely that students' language level plays a role. Therefore, the role of language is an important point of focus in the current study.

Many older studies were carried out in the field of word problems, particularly in the domain of addition and subtraction with young children (e.g., see Carpenter, Moser, & Romberg, 1982 and for a recent overview see Verschaffel, Greer, & De Corte, 2007). *Word problems* can be considered a subcategory of the broader class of mathematics problems presented in a realistic context. This word problem research has shown that problem solving behavior is influenced by characteristics of the students (e.g., how advanced they are) and characteristics of the problems (e.g., particular semantic features), and the interaction between the two. Therefore, it is important to specify the population of students and type of

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

problems addressed. The major objective of the current study was to investigate students and problems typical for the regular school setting, in order to obtain high ecological validity of the conclusions and implications for educational and testing practices. To that end, contextual multidigit arithmetic problems involving one of the four basic operations (addition, subtraction, multiplication, and division) were selected from national assessment studies. These problems, involving one-step arithmetic operations, were considered typical for contextual problems in school mathematics. Regarding the population of students, we focused on students at the end of primary school (i.e., sixth grade) for at least two reasons. First, by investigating the influence of contexts at a time point where students have received many years of formal schooling in solving mathematics problems with and without a context, we aimed to extend existing research mainly focusing on young children. Second, students at the end of primary school are often subjected to high-stakes educational tests, of which the mathematics subtests increasingly rely on problems presented in a context. Therefore, it is important to investigate the possible effects of such contexts in mathematics tests specifically for this group of students.

Solving numerical and contextual arithmetic problems

Solving numerical and contextual problems (sometimes referred to as 'computations' and 'applications', respectively) is likely to involve different aspects of mathematical cognition. At least two different perspectives exist. In one perspective, it is stressed that solving contextual problems involves a complex process consisting of several cognitive processes or phases (e.g., Fuchs et al., 2006, 2008; Wu & Adams, 2006). Specifically, first an accurate situational and mathematical model of the problem situation has to be formed (a process called mathematization), before computational skill – and carefulness therein – comes into play. Therefore, factors other than 'pure' computational skills are likely to contribute to success in solving contextual problems. This perspective yields the expectation that contextual problems are more difficult to solve than numerical problems, as supported by early research findings of Cummins et al. (1988) in

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

simple addition and subtraction word problems.

By contrast, in the other perspective on the effect of problem representation, it is argued that contextual problems can also be easier to solve than numerical problems. That is, the realistic context may activate real-world knowledge which can aid problem-solving by eliciting other strategic approaches (e.g. Gravemeijer, 1997). In the domain of algebra, Koedinger and Nathan (2004) and Koedinger, Alibali, and Nathan (2008) found that students were more successful in solving so-called grounded representations (story problems) than in solving algebra equations presented symbolically, because they used intuitive, informal strategies more often to solve the story problems. However, this advantage of story problems held only for simple problems and not for more complex ones, a pattern the authors called the representation-complexity trade-off. They argued that both grounded and abstract representations each have specific advantages based on their different properties: grounded representations are expected to be more familiar and less error-prone, while abstract representation are more concise and put fewer demands on working memory (Koedinger et al., 2008).

Both perspectives have in common, however, that there are at least two ways in which the problem format (contextual versus numerical) may affect performance: different ability dimensions may be involved in solving problems with and without a realistic context, or the context may affect the difficulty level of a problem.

Recently, a few studies empirically investigated to what extent different abilities were involved in solving mathematic problems of both formats in American third graders (Fuchs et al., 2006, 2008) and in Dutch first to third graders (Hickendorff, in press). These studies showed that solving numerical mathematics problems and solving contextual problems involved two highly related but distinct ability dimensions, as evidenced by a less than perfect correlation between performance measures and by different cognitive correlates for the two measures. However, these studies did not allow a direct investigation of how a realistic context affects the difficulty level of a problem, because the numerical characteristics of contextual and numerical problems were not matched. In contrast, Vermeer,

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Boekaerts, and Seegers' (2000) study on sixth graders' arithmetic problem solving involved matched computation (i.e., numerical) problems and application (i.e., contextual) problems, thereby potentially allowing for direct comparisons. Regrettably, direct tests comparing performance on the two types of problems were not reported, although the proportion correct was found to be slightly higher on the application problems than those involving computation. Since theoretical hypotheses and empirical results are inconclusive on the effects of problem format on performance, further systematic study is needed.

The present study extends the previously discussed studies in three ways. First, the effect of problem format (with or without a context) on problem solving is investigated in a systematic test design, consisting of problem pairs in which one problem was presented with a realistic context and the matched parallel problem without such a context. Second, a more complete account of problem solving was taken: not only performance but also strategy use is researched. Finally, given the verbal nature of contextual problems, the role of students' language level is addressed.

Solution strategies

Performance or accuracy is probably the most salient aspect of problem solving, and many studies into solving problems with and without a realistic context focused only on that aspect (Fuchs et al., 2006, 2008; Hickendorff, in press; Vermeer et al., 2000). However, another important aspect of problem solving is strategic competence. In cognitive psychology it is well-established that adults and children know and use multiple solution strategies to solve mathematics problems (e.g., Lemaire & Siegler; 1995 Siegler, 1988a). Furthermore, solution strategies are also important from the perspective of mathematics education in at least two ways. First, the didactics for solving complex arithmetic problems have changed, from instructing standard written algorithms to building on children's informal or naive strategies (Freudenthal, 1973; Treffers, 1987, 1993), and mental arithmetic has become very important (Blöte, Van der Burg, & Klein, 2001). Second, mathematics

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

education reform aims at attaining adaptive expertise instead of routine expertise: instruction should foster the ability to solve mathematics problems efficiently, creatively, and flexibly, with a diversity of strategies (Baroody & Dowker, 2003; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009).

Lemaire & Siegler (1995) distinguished four aspects of strategic competence: strategy repertoire, strategy choice, strategy performance (such as accuracy), and strategy adaptivity. The current study focuses on the first three of these aspects, on the domain of multidigit arithmetic. In the domain of elementary or simple arithmetic, strategy use has been studied extensively: in elementary addition and subtraction (e.g., Carr & Jessup, 1997; Carr & Davis, 2001; Torbeyns, Verschaffel, & Ghesquière, 2004, 2005), in elementary multiplication (e.g., Anghileri, 1989; Imbo & Vandierendonck, 2007; Lemaire & Siegler, 1995; Mabbott & Bisanz, 2003; Mulligan & Mitchelmore, 1997; Sherin & Fuson, 2005; Siegler, 1988b), and in elementary division (e.g., Robinson et al., 2006). By contrast, research on solution strategies in complex or multidigit arithmetic problems is less extensive, but there is a growing body of studies in multidigit addition and subtraction (e.g., Beishuizen, 1993; Beishuizen, Van Putten, & Van Mulken, 1997; Blöte et al., 2001; Torbeyns, Verschaffel, & Ghesquière, 2006) and in multidigit multiplication and division (e.g., Ambrose, Baek, & Carpenter, 2003; Buijs, 2008; Hickendorff, Heiser, Van Putten, & Verhelst, 2009; Hickendorff & Van Putten, 2012; Hickendorff, Van Putten, Verhelst, & Heiser, 2010; Van Putten, Van den Brom-Snijders, & Beishuizen, 2005).

The current study addressed multidigit arithmetic involving the four basic operations. Based on the solution strategies reported in the aforementioned studies in multidigit addition, subtraction, multiplication, and division (see also a recent review by Verschaffel et al., 2007), a classification scheme of written solution strategies was developed (i.e., the strategy repertoire). For each of the four operations, a basic distinction can be made (see also Selter, 2001) between the *traditional standard algorithm* that proceeds digit-wise, *non-traditional procedures* that are more informal, dealing with whole numbers, and *answers without written working*, most likely mental strategies. A subcategory of the non-traditional

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

strategies are the RME approaches (labeled 'columnwise arithmetic' by the developers, see Van den Heuvel-Panhuizen, Buys, & Treffers, 2001, and Treffers, 1987). These can be considered transitory between informal approaches and the traditional algorithm: they work with whole numbers instead of single-digits (like informal strategies), but they proceed in a more or less standard way (like the written algorithm). More details are given in the Method-section.

Based on the literature, we had the following expectations regarding the effects of problem format (contextual or numerical) on strategy use. Studies on elementary word problem solving with young children showed that different semantic structures of word problems elicited different strategies (for a review, see Verschaffel et al, 2007). Extending these findings leads to the expectation that contextual and numerical problems on multidigit arithmetic elicit different strategies. In particular, the theory behind the RME didactical approach yields the expectation that problems in a realistic context are more likely to elicit more informal, less structured strategies (i.e., non-traditional or mental solution strategies), while numerical problems would elicit more use of traditional algorithms, as hypothesized by Van den Heuvel-Panhuizen, Robitzsch, Treffers, & Köller (2009). This pattern was indeed found in Koedinger and Nathan's (2004) and Koedinger et al.'s (2008) study on algebra problem solving. In contrast, Van Putten et al. (2005) investigated Dutch fourth graders strategy use on multidigit division problems that either did or did not include a context, and found no differences in strategy choice between the two types of problems. Given these inconsistent findings, the effects of contexts on strategy use require further systematic study.

The role of students' language level and gender

Because it is necessary that students accurately understand the usually verbal problem situation of a contextual problem, it is likely that the student's language ability plays an important role. Support for the importance of language in word problem solving comes from the finding that language ability had larger effects on applied problem solving (contextual problem solving) than on computational skills (numerical problems) (Fuchs et al., 2006, 2008; Hickendorff, in

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

press). Additional support comes from the finding that a common source of errors in word problem solving appears to be misunderstanding of the problem situation (Cummins et al., 1988; Wu & Adams, 2006), and that conceptual rewording of word problems facilitated performance (e.g., Vicente, Orrantia, & Verschaffel, 2007). Therefore, we expect the effect of language ability level to be larger on performance in solving contextual problems than in solving numerical problems.

Ethnic minority students score lower on language ability tests than native students. In addition, international assessments such as TIMSS-2007 (*Trends in International Mathematics and Science Study*; Mullis, Martin, & Foy, 2008) and Dutch national assessments (J. Janssen et al., 2005; Hop, 2012) consistently report that ethnic minority pupils lag behind in mathematics achievement too. An obvious question is whether language level plays a role in the performance lag of ethnic minorities on mathematics problems that involve a verbal context. Several research findings with students in secondary education showed that the difficulty of the problem text particularly hampers non-native speakers (Abedi & Hejri, 2004; Abedi & Lord, 2001; Prenger, 2005; Van den Boer, 2003), due to text aspects like the use of unfamiliar vocabulary, passive voice construction, and linguistic complexity. Therefore, we expect differences with respect to the language spoken at home to be larger on the performance in solving contextual problems than in solving numerical problems.

A final student characteristic addressed in the present study is gender. Gender differences in general mathematics performance have been reported frequently. Large-scale international assessments TIMSS-2007 (Mullis et al., 2008) and PISA-2009 (OECD, 2010) showed that boys tend to outperform girls in most of the participating countries, including the Netherlands. This pattern is supported by Dutch national assessments findings: on most mathematical domains boys outperformed girls in third and in sixth grade (J. Janssen et al., 2005; Hop, 2012). However, in grade 6, the multidigit operations were the exception with girls slightly outperforming boys. Moreover, Vermeer et al. (2000) found that in Dutch sixth graders, there were no gender differences in performance on computations, while

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

boys outperformed girls on applications. The latter difference may possibly be explained by the finding that on application problems, girls had lower levels of subjective competence than boys and attributed bad results to lack of capacity and difficulty of the task. Based on these results, we expect that gender differences in performance to be larger on contextual problems than on numerical problems.

Regarding strategy choice, girls have been found to be more inclined to (quite consistently) rely on rules and procedures and use well-structured strategies, whereas boys have a larger tendency to use more intuitive strategies (Carr & Davis, 2001; Carr & Jessup, 1997; Gallagher et al., 2000; Hickendorff et al., 2009, 2010; Hickendorff & Van Putten, 2012; Timmermans, Van Lieshout, & Verhoeven, 2007; Vermeer et al., 2000). There are no empirical findings on whether this pattern is the same for numerical problems as for contextual problems, so further study is needed.

The current study

The current study's aim was to systematically investigate the effects of presence of a context in mathematics problems on two aspects of problem solving: performance and strategy use (strategy choice and strategy accuracy). To investigate this issue with students and problems typical for the regular school setting, contextual multidigit arithmetic problems on the four basic operations (addition, subtraction, multiplication, and division) were selected from national assessment tests. These problems were administered to a sample of sixth graders in two formats: once with the context, and once without (i.e., in bare numerical format). The total problem set consisted of eight such pairs of problems.

Based on the previous discussion of existing theoretical literature and empirical findings, we had the following expectations. With regard to performance, we expected that two highly related but distinct ability dimensions would be involved in solving the two types of problems, and that contextual problems are more difficult, in particular for students with low language level as well as for girls. With regard to strategy use, we expected that contextual problems would elicit more use of informal, less structured strategies than numerical problems.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Method*Participants*

Participants were 685 students from grade 6 with mean age 12 years 0 months (SD = 5 months). They originated from 24 different primary schools, with 3 to 82 students participating per school (on average 27.4 students per school; all sixth graders in a class participated). These schools were spread over the entire country of the Netherlands. There were 312 boys, 337 girls, and 36 students with missing gender information. In order to assess language level effects with sufficient power, the schools that were selected had relatively high ethnic minority populations. As a consequence, the current sample of schools and pupils was not entirely representative for the population of Dutch primary schools.

Information on the language spoken at the students' home was gathered (missing data for 42 students). Students were classified into home language *Dutch* (either only Dutch, 517 students, 80%; or Dutch as well as another language, 46 students, 7%) or home language *non-Dutch* (80 students, 12%). The most prevalent non-Dutch language was Arabic (45% of students with home language other than Dutch), followed by Turkish (26%).

Material

Experimental task. The experimental task consisted of 16 multidigit arithmetic problems, consisting of 8 pairs of one contextual and one numerical problem each. There were 2 pairs of problems for each multidigit operation: addition, subtraction, multiplication, and division (with whole number outcomes), see Appendix A.

The contextual problems were selected from the most recent Dutch national assessment (J. Janssen et al., 2005). Two problems were selected for each operation: one at the lower end of the ability scale (i.e., a relatively easy problem with small numbers), and one at the upper end of the scale (i.e., a relatively hard problem with

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

large numbers). We used problems from the assessments to ensure that they were representative for the type of contextual problems that are used in current educational practices. For the numerical problems, these contextual problems were disposed of their contexts to yield the bare numerical operation required. In this process, decisions had to be made on the order of the numbers in addition and multiplication, and on whether to use a canonical or non-canonical form in subtraction and division. For instance, the first contextual division problem from Appendix A can have different numerical variants, such as $736 : 32 = ?$ (the one chosen in this study) or $32 \times ? = 736$. The choice for a particular numerical problem was made based on the most likely approaches of students to solving typical contextual problems, resulting from earlier studies (Hickendorff et al., 2009; Hickendorff & Van Putten, 2012). Furthermore, in order to avoid testing effects that may have occurred if students had to solve exactly the same numerical operation twice (once with and once without a context), a parallel version of each problem was constructed with numbers and solution steps as similar as possible.

Two different test forms were created, so that the item parallel version was counterbalanced over the test form. That is, in form A item versions *a* were presented as contextual problems and item versions *b* as numerical problems, and in form B this pattern was reversed. For example, the first item pair on Addition in Appendix A presents the problems as presented in form A. In form B, the numbers were switched: i.e., the text of the contextual problem said that 677.50 euro was sold on postcards and 975 euro on stamps, while the numerical problem was $466.50 + 985 = ?$. Figure 1 presents the specific position of each problem in both task forms. Within each form, paired problems (e.g. A_{1a} and A_{1b}) were presented with 7 other problems in between, to prevent recency effects. The order of the 16 different problems was the same in both task forms, to rule out potentially confounding order effects in combining the data from the two forms.

In the test booklets that students received, a maximum of 3 problems were printed on the left side of a page (A4 size). The right side of each page was left blank, so that students could use that space as scrap paper in solving the problems.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Standardized tests. The students participating in the current study took part in the 2009 administration of CITO's End of Primary School Test (CITO, 2009) in February. This test is widely used in the Netherlands at the end of primary school, and its purpose is to give advice on the most suitable track of secondary education for each student. To that end, the instrument assesses scholastic achievement level in mathematics, language, and study skills. Over 150,000 Dutch sixth graders participated in the 2009 assessment. The 100-item subtest on language skills consisted of items on writing, spelling, reading comprehension, and vocabulary, and had high internal reliability ($KR20 = .89$; CITO, 2009). In the current sample, the average number of language items correct was 73.0 ($SD = 11.7$; missing data for 29 students). This mean score was slightly lower than for the entire population of students participating in the End of Primary School Test, who scored on average 75.2 items correct ($SD = 12.0$). On the 60-item subtest on mathematics ($KR20 = .91$), the current sample scored on average 41.7 items correct ($SD = 10.6$), which was also slightly lower than the average of 42.8 correct ($SD = 10.5$) for all participants nationwide.

Procedure

Experimental task. The experimental task was administered as part of a pretest study for the CITO End of Primary School Test. A test booklet consisted of 6 tasks, divided over the different subjects mathematics, language, and study skills. Students completed each of these tasks on a separate day in January 2009. One of the mathematics tasks included the current experimental task of 16 problems, and an additional 12 problems that were not part of the current study. One of the two experimental task forms (A or B) was assigned to each class.

The task was administered in the classroom, and each student worked individually. Teachers instructed their students that they were free to choose their solution strategy. Students were also told that they could use the blank space next to each problem in the test booklet to make computations, and that they did not need separate scrap paper. Students could take as much time as they needed, so there was no time pressure.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Standardized tests. The students completed the 2009 End of Primary School Test (CITO, 2009) as part of their final year’s standardized assessment in February 2009, which was at most one month after the students participated in the current study.

Solution strategies

The solution strategy used on each trial (student-by-item combination) was categorized based on the notes or solution procedures that students had written in the test booklet. Strategy data were available for 650 students. Three experts (the first author and two trained research assistants) each coded a separate part of the material. Table 1 shows the 9 (addition) or 10 (subtraction, multiplication, and division) different categories of solution strategies that were distinguished, and Appendix B shows examples of categories 1 to 5 for each operation. Below, first the operation-specific categories 1 to 6 are discussed, and after that the operation-general categories 7 to 10 are explained.

Addition. Category 1 (traditional algorithm) coded strategies in which the standard algorithm for adding two or more multidigit numbers was applied. The addends have to be aligned vertically so that digits on the same position represent the same value, and addition proceeds digit-wise from right to left starting with the ones-digits (assuming there are no decimals), next the tens-digits, then the hundreds-digits, and so forth. If the outcome of any particular sub-addition is larger than 10, regrouping has to take place. Category 2 is the RME approach to addition. It contrasts with the traditional algorithm because it proceeds from left to right and it works with numbers instead of single-digits (e.g., $600 + 900 = 1500$ instead of $6 + 9 = 15$). Category 3 and 4 are partitioning strategies, in which either one or more than one of the operands is partitioned or split according to its place value (e.g., 975 is split into 900, 70, and 5). Partitioning of only the second operand is also called the jump or sequential strategy, while partitioning of two or more operands is called the split or decomposition strategy (e.g., Beishuizen, 1993). The final category 6 (note that we left out category number 5 for addition to be consistent with the other

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

operations) included all kinds of strategies in which some calculations or intermediate solutions were written down from which could be inferred how the answer was obtained, but that did not fit in categories 1 to 4.

Subtraction. Category 1 (traditional algorithm) involved application of the standard algorithm for subtraction of two multidigit numbers. Similar to the addition algorithms, the two numbers have to be aligned vertically so that digits on the same position represent the same value, and subtraction proceeds digit-wise from right to left starting with the ones-digits (assuming there are no decimals), next the tens-digits, then the hundreds-digits, and so forth. In case a larger digit has to be subtracted from a smaller one (e.g., 0 - 9), regrouping has to take place. Category 2 is the RME approach to multidigit subtraction. It contrasts with the traditional algorithm because it proceeds from left to right and it works with numbers instead of single-digits. Moreover, there is no need of borrowing: it works with negative numbers instead (e.g., $10 - 80 = -70$). Category 3 and 4 are partitioning strategies, in which either only the subtrahend or both operands are partitioned according to their place value (e.g., 689 is split into 600, 80, and 9). Similar to addition, partitioning of only the subtrahend is also called the jump or sequential strategy, while partitioning of both operands is called the split or decomposition strategy (Beishuizen, 1993). Category 5 involved indirect addition strategies. In these approaches, one starts from the subtrahend and adds on until the minuend is reached (see for example Torbeyns, Ghesquière, & Verschaffel, 2009). The final category 6 (other written strategy) included all kinds of strategies in which some calculations were written down, but that did not fit in categories 1 to 5.

Multiplication. The traditional standard algorithm for multiplication (category 1) involves writing the two operands below each other, and multiplying the upper number by each digit of the lower number separately, working from right to left. Then, these partial outcomes are added to obtain the solution. In the RME approach (category 2), both numbers are partitioned and all sub-products are obtained and added. This strategy closely resembles the one in category 4 (partitioning of both operands), but the difference lies in the schematic notation that is applied in category 2 which is stressed in the RME approach. Hence, these two

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

categories are separated because their distinction is educationally relevant more than because it is of interest from a cognitive perspective. In category 3, only one of the operands is partitioned, while the other one is left intact (e.g., $36 \times 27 = 36 \times 20 + 36 \times 7$). Category 5 involved repeated addition, in which there is made use of the fact that multiplying by a factor n is equivalent to adding the multiplicand n times. This category included strategies in which either the multiplicand was added n times, or when doubling strategies were used. Again, the final category 6 included all kinds of strategies in which some calculations were written down, but that did not fit in categories 1 to 5.

Division. The first category of division was the long division algorithm (note that notation may differ between countries). The algorithm is characterized by starting on the left side with places of highest value of the dividend, and trying to divide the first digit by the divisor (e.g., $7 \div 32$). If that yields a number smaller than 1, the first two digits are considered together (73) and the maximum number of times the divisor fits in ($2 \times 32 = 64$) is noted. Then, the difference is determined ($73 - 64 = 9$) and the digit from the column to the right is pulled down (making 96). This procedure continues until all digits of the dividend have passed and the remainder is zero, or in case of a nonzero remainder until the required level of precision in the quotient is reached. Categories 2 and 3 involve repeated subtraction strategies (in the Netherlands this is the RME alternative for long division in the mathematics textbooks, Treffers, 1987). Multiples of the divisor are repeatedly subtracted from the dividend. This can be done efficiently with relatively few steps (high-level, HL) or less efficiently with many steps (low-level, LL). In the present study, we defined strategies as high-level when at most 3 steps (the minimum number of steps + 1) were taken. It is worth noting that the most efficient repeated subtraction strategy resembles the traditional algorithm, with the main difference that in the algorithm one ignores the place value of the digits, while one works with whole numbers in repeated subtraction (e.g., with 640 instead of 64). Categories 4 and 5 resemble categories 2 and 3, respectively, but they differ in the approach: repeatedly adding multiples of the divisor until the dividend is reached, as opposed to repeatedly

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

subtracting from the dividend until zero is reached. The same distinction between high-level (maximum 3 steps) and low-level approaches was made. As with the other operations, category 6 (other written strategy) involved strategies in which some calculations were written down, but that could not be classified in categories 1 to 5.

Remainder categories. The remaining strategy categories 7 to 10 were the same for the four operations. Category 7 (no written working) includes all trials in which an answer was written down, but nothing else (i.e., no calculations or intermediate solutions), so it is very likely that the answer was computed mentally (supported by findings of Hickendorff et al., 2010). Category 8 (wrong procedure) includes trials in which the wrong procedure was applied, such as adding the two numbers in a division problem. In trials classified in category 9 (unclear strategy) it was unclear how the student arrived at the answer (s)he had given, in some cases because the written solution steps were erased. The final category 10 (skipped problem) included trials in which the problem was skipped entirely, i.e., no answer was given and no solution steps were written down.

Reliability. The solution strategies of 45 students (720 trials; 180 trials per operation) were double-coded by two independent raters to assess the agreement in categorization. Cohen's kappa (Cohen, 1960) on the cross-tabulation of the categorization of the two raters was computed as a measure of inter-rater reliability. Kappa values were sufficiently high with .82, .79, .89, and .92 for addition, subtraction, multiplication, and division, respectively, indicating substantial and satisfactory agreement.

Data Analysis and Results

In all data-analyses, the data were collapsed over the two test forms A and B, thereby counterbalancing potential differences between parallel item versions within problem pairs. The results are presented in two parts: performance and strategy use.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Performance

The proportion correct on contextual problems and numerical problems were very close on addition (.70 vs. .72), subtraction (.72 vs. .73), multiplication (.69 vs. .68), and division (.75 vs. .73). Recall that the main question was: What is the effect of presence of a context on performance?, and that this question was addressed in two ways. First, we established whether separate dimensions of individual differences were involved in solving items with and without a context (the multidimensionality hypothesis). Second, the effect of problem format on the difficulty level of an item was tested (the sources-of-difficulty hypothesis).

Because the accuracy data concern repeated dichotomous observations (i.e., each student answers 16 items), analysis techniques that can take this multivariate character into account are needed. Aggregating over items by computing total proportion correct, as is characteristic of the Classical Test Theory (CTT) approach, has several disadvantages. One salient disadvantage is that the resulting proportion correct scores are potentially hampered by ceiling and floor effects, biasing subsequent statistical analyses on these scores. Modern test theory, specifically item response theory (IRT) modeling, overcomes these limitations by accounting for the influence of item characteristics on the responses (for a detailed discussion of the advantages of IRT over CTT, see Embretson & Reise, 2000). In IRT models, one or more continuous latent variables are introduced that model individual differences in ability level, affecting the probability to solve a particular item correctly. Specifically, in the most simple unidimensional IRT model – the Rasch model – the probability P_{is} that person s solves item i correctly is modeled as a logistic function of the difference between the person’s latent ability level θ_s and the item’s difficulty level

$$\beta_i: P_{is} = \frac{\exp(\theta_s - \beta_i)}{1 + \exp(\theta_s - \beta_i)}.$$

Both hypotheses were addressed with item response theory (IRT) models. We used the `lmer` function from the `lme4` package (Bates & Maechler, 2010) available in the statistical computing program R (R Development Core Team, 2009) to estimate the model parameters. For further details on how to fit IRT models with

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

lmer, see De Boeck et al. (2011). The main assumptions of IRT models are that the model involves the correct number of dimensions of individual differences and that the item response curves P_{is} follow the specified mathematical function (Embretson & Reise, 2000). The first assumption about dimensionality is at the core of the current investigation. The second assumption will be tested by checking for item fit statistics that quantify how well the observed item responses are predicted by the estimated IRT model parameters (Embretson & Reise, 2000).

Multidimensionality. To explore whether different latent abilities were involved in solving items with and without a context, we used multidimensional IRT (MIRT) modeling (see Reckase, 2009). Specifically, a confirmatory two-dimensional IRT model was used, in which each item was assigned a priori to one of the two dimensions *solving numerical problems* and *solving contextual problems*. Figure 2 shows a graphical display of this model. Such a model belongs to the class of between-item or simple structure Rasch models, in which it is assumed that multiple related subscales or ability dimensions underlie test performance, and that each item in the test is only related to one of these subscales (Adams, Wilson, & Wang, 1997).

Our main interest was in the estimate of the latent correlations between the two ability dimensions θ_{num} and θ_{con} . A latent correlation estimate in a MIRT model is not attenuated by measurement error: it is an unbiased estimate of the true correlation between the latent variables (Adams & Wu, 2000; Wu & Adams, 2006). Therefore, it is a better alternative than computing the correlation on ability estimates of consecutive unidimensional models, or on classical test theory approaches that are based on the proportion of items solved correctly (as was done in the studies by Fuchs et al., 2006, 2008).

The results of fitting a 2-dimensional between-item Rasch model showed that the latent correlation between the ability to solve numerical problems and the ability to solve contextual problems was estimated at 1.000.² Therefore, we conclude that solving numerical problems and solving mathematics problems in a

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

context involves one (latent) ability factor. Furthermore, we computed item fit statistics for the unidimensional Rasch model with the `ltm` package (Rizopoulos, 2006). There were no items with significant misfit (lowest p -value was .059), indicating that the unidimensional Rasch model adequately predicted the observed item response patterns.

Sources of item difficulty. The next step was to assess the effect of item format on the difficulty level of an item. For that end, we used an extended Linear Logistic Test Model (LLTM; Fischer, 1987). The LLTM is an example of a broad class of explanatory IRT models, in which predictors on the item level, the person level, and on the item-by-person level can be incorporated in the model (De Boeck & Wilson, 2004). The LLTM allows for decomposition of item difficulty β_i into the effects of K different item features, in a multiple regression-like manner: $\beta_i = \sum_{k=1}^K \tau_k q_{ik} + \tau_0$. The q_{ik} entries of the so-called Q-matrix specify the involvement of item feature k in item i , and have to be assigned a priori. The LLTM has the drawback that it assumes that the K item features predict item difficulty without error. To relax this assumption, the LLTM can be extended by incorporating error that is randomly distributed over items in the model, yielding the LLTM + e model (De Boeck, 2008; R. Janssen, Schepers, & Peres, 2004).

There were three item features: operation required (nominal variable with 4 categories: addition, subtraction, multiplication, and division; recoded into 3 dummy variables), number size of the problem (dichotomous variable with 2 categories: small or large), and item format (dichotomous variable with 2 categories: numerical or contextual). Our main interest was in the effect of item format, statistically correcting for the covariates operation and number size. In addition, we corrected for possible group differences in ability level between students who were administered test form A or test form B, by including 'test form' in the model as well. Because this was a variable on the student level, the final IRT models that we used could be characterized as *latent regression LLTM + e* models.

Results showed that the main effect of item format was not significant ($\tau_{\text{context}} = .05, p = .30$). Testing for interactions between item format and the other two item

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

features, it was found that the interaction between item format and number size was not significant (Likelihood Ratio $\chi^2(1) = .0$, $p = 1.00$), while the interaction between item format and operation was significant ($\chi^2(3) = 16.4$, $p < .001$). Further inspection of this latter interaction showed that the effect of context format was significant only on division problems, $\tau_{\text{context, division}} = .35$, $p < .001$, while for the other three operations the effect of context format was not significant ($\tau_{\text{context, addition}} = -.15$, $p = .10$; $\tau_{\text{context, subtraction}} = -.08$, $p = .36$, and $\tau_{\text{context, multiplication}} = .08$, $p = .35$). So, only for division problems the context made the item more difficult compared to the bare numerical problem. On the other three operations the contextual problems and the numerical problems were just as difficult.

Final analyses were carried out to assess whether the effect of item format (numerical vs. contextual) on item difficulty depended on gender, language achievement level, or home language. First, the main effects of the student characteristics on performance are reported. Gender had a significant effect on performance: girls had significantly more correct answers than boys ($\beta_{\text{girl}} = .23$, $p = .03$). In contrast, the effect of home language on performance was not significant ($\beta_{\text{other than Dutch}} = -.04$, $p = .82$). Furthermore, the effect of language achievement level was highly significant and positive: $\beta_{\text{language}} = .58$, $p < .001$. Finally, we focus on the interactions between student characteristics and item format. The interaction between item format and gender ($\chi^2(4) = 4.4$, $p = .35$), home language ($\chi^2(4) = 1.1$, $p = .89$), and language achievement level ($\chi^2(4) = 4.1$, $p = .39$) all turned out to be nonsignificant. So, the effects of presenting an item in a context did not depend on either students' gender, home language or language achievement level.

Strategy use

Strategy choice. Table 2 shows the distribution of the strategy categories for addition, subtraction, multiplication, and division, split by problem format (numerical versus contextual). There were only small differences with respect to problem format in the distribution of strategies: The percentages of trials solved by each strategy were very similar for the numerical problems as for the contextual problems, with the largest difference being 3 percent points.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

In contrast, operation required seemed to have large effects on strategy choice distribution. First, differences in the use of the traditional algorithm (category 1) emerged. It was the dominant strategy for addition and subtraction (used on around 70% of the trials). Also for multiplication it was the most prevalent strategy although its dominance was less pronounced, being used on over 50% of the multiplication trials. For division, however, the traditional algorithm was used on only 15% of all division trials. Second, the frequency of answering without written working (category 7) was about the same for each operation, occurring on between 14% and 17% of all trials. Moreover, wrong procedures (category 8), unclear strategies (category 9), and skipping the entire problem (category 10) did not occur very often. These occurred with about the same frequency on each operation, although slightly more often on division. Third, some operation-specific patterns emerged. Most notably, on division the high-level repeated subtraction strategy was the most prevalent strategy, used on over 50% of all division trials. On multiplication, partitioning of 1 operand (category 3) was used quite often (17% of all trials). Finally, on addition, the RME approach was used relatively often (9% of all trials) compared to the other operations.

In order to statistically test the effects of operation and problem format on strategy choice distribution, we first collapsed some of the categories to make strategy categories comparable across operations, and also to obtain categories filled with a substantial number of observations. We recoded the 9 or 10 operation-specific strategies into 4 operation-general categories: the traditional algorithm (former category 1), non-traditional strategies (former categories 2 to 6), no written working (former category 7), and other trials (former categories 8 to 10). Figure 3 presents the proportion of choice of each of these four strategy types on numerical and contextual problems, per operation.

Next, we estimated a multinomial logistic model for correlated responses using a random effects model (Hartzel, Agresti, & Caffo, 2001) in the SAS procedure NLMIXED, as described by Kuss & McLerran (2007). The main assumption of this model is that the random effects follow a multivariate normal distribution, which is a common assumption in mixed effect models (such as multilevel analysis) that is

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

hard to test. As predictor variables in the model, we first included only variables on the item level: operation required (nominal variable with 4 categories: addition, subtraction, multiplication, or division; recoded into 3 dummy variables) and problem format (numerical or contextual), and their interaction.

Operation had a highly significant effect (Likelihood Ratio $\chi^2 (9) = 3995.4, p < .001$) on the distribution of strategies, as is also obvious from Figure 3. However, the main effect of problem format was not significant ($\chi^2 (3) = 4.9, p = .18$), and neither was the interaction between problem format and operation ($\chi^2 (9) = 15.8, p = .07$). Therefore, we may conclude that the presence of a context did not influence the distribution of solution strategies on any of the four operations.

Final analyses were carried out to assess whether the effect of item format (numerical vs. contextual) depended on gender, language achievement level, or home language. First, the main effects of the student characteristics on strategy choice distribution are reported. The effect of home language was not significant ($\chi^2 (3) = .4, p = .95$), so students who spoke Dutch at home had the same strategy distribution as students who spoke another language at home. In contrast, gender ($\chi^2 (3) = 36.4, p < .001$) as well as language achievement level ($\chi^2 (3) = 187.7, p < .001$) had a significant effect on strategy distribution. Table 3 shows that girls were more likely than boys to choose the traditional algorithm, and less likely to answer without written working. To visualize the effect of language achievement level, it was recoded into three categories based on the population percentile rank (based on all participants of the End of Primary School Test 2009): low (up to percentile 33), medium (percentiles 34 to 66), and high (percentile 67 and higher). Table 3 shows that the probability to choose the traditional algorithm increased with higher language level, while the probability of answering without written working as well as choosing one of the other strategies decreased with higher language level.

Finally, the interaction effects between problem format (numerical vs. contextual) on the one hand, and the student characteristics on the other hand, were not significant regarding gender ($\chi^2 (3) = .4, p = .95$), home language ($\chi^2 (3) = .9, p = .82$), or language achievement level ($\chi^2 (3) = 2.7, p = .44$). Thus, we may conclude

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

that the finding that the presence of a context did not affect strategy choice distribution holds for all subgroups of students.

Strategy accuracy. Finally, we investigated to what extent the problem format (numerical vs. contextual) affected the accuracy of each of the strategies, i.e., the proportion correct per strategy. To that end, we again used explanatory IRT models (De Boeck & Wilson, 2004). This time, not only predictors on the item level were included, but also the strategy used was included as a person-by-item predictor (see also Hickendorff et al., 2009, 2010). All trials in which the solution strategy was classified in the Other category were excluded from the analyses, because this was a small heterogeneous group of trials with many skipped problems. As a result, we analyzed the accuracy differences between the traditional algorithm, non-traditional strategies, and no written working.

Results showed that these three strategies differed significantly in accuracy ($\chi^2 (2) = 184.6, p < .001$), and that the accuracy differences depended on the operation required ($\chi^2 (6) = 55.0, p < .001$). However, the operation-specific accuracy differences between the strategies did not depend on the item format (numerical vs. contextual), $\chi^2 (8) = 9.0, p = .34$. Figure 4 shows the estimated proportion correct of each strategy for students at the mean of the latent ability scale, by operation.

The following pattern emerged: the traditional algorithm had higher accuracy than the non-traditional strategies, which in turn were more accurate than no written working. Statistical testing of these differences showed the following: the regression parameters of the accuracy difference between the traditional algorithm and non-traditional strategies were $\beta = .32 (p = .02)$, $\beta = .96 (p < .001)$, $\beta = .55 (p < .001)$, and $\beta = .10 (p = .60)$, for addition, subtraction, multiplication, and division, respectively. So, on addition, subtraction, and multiplication the traditional algorithm was significantly more accurate than the non-traditional strategies, while the difference was not significant on division. Next, the regression parameters for the accuracy difference between non-traditional strategies and no written working were $\beta = .58 (p < .001)$, $\beta = .09 (p = .67)$, $\beta = .74 (p < .001)$, and $\beta = 1.72 (p < .001)$, for

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

addition, subtraction, multiplication, and division, respectively. Thus, on addition, multiplication, and division, non-traditional strategies were significantly more accurate than no written working, while the difference was not significant on subtraction. Moreover, it is worth noticing that the estimated accuracy of no written working on division problems was much lower than on the other three operations.

Discussion

The current study aimed to assess the effects of presenting multidigit arithmetic problems in a realistic context typical for school mathematics, on two aspects of sixth graders' problem solving: performance and solution strategy use. First, regarding performance, multidimensional IRT models showed that the same latent ability dimension was involved in solving numerical and contextual problems. Moreover, explanatory IRT models showed that presenting an arithmetic problem in a context increased the difficulty level of the division problems, but did not affect the difficulty level of addition, subtraction, and multiplication problems. Importantly, these performance effects were independent of students' gender, home language, and language achievement level. Second, the presence of a context did not affect the strategy choice distribution, nor the strategy accuracy, irrespective of students' gender, home language, and language achievement level. In summary, we conclude that, contrary to our expectations, the effects of presenting arithmetic problems in a realistic context typical for school mathematics tests were nonexistent in addition, subtraction, and multiplication, and that only a small difference in problem difficulty was found for division.

Regarding performance, based on earlier research findings (Fuchs et al., 2006, 2008; Hickendorff, in press) we expected that related but separate ability dimensions would be involved in solving the two types of problems. A possible explanation for the difference between the current results and previous findings may lie in the differences between the age groups (first to third graders versus sixth graders). It has been argued that children in higher grades, who have had more years of formal schooling, have more developed cognitive schemata to solve word

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

problems (De Corte et al., 1985; Vicente et al., 2007). Possibly, these cognitive schemata are so well-developed at grade 6 that students no longer perceive differences between such typical contextual problems and numerical problems. This would result both in indistinguishability of the latent ability dimensions involved, as well as in absence of an effect on problem difficulty (as was also found in another study in sixth grade by Vermeer et al., 2000). Importantly, this pattern held for girls, for students with low language ability level, and for students from non-native origin. It thus seems that, at the end of primary school, these students are not hampered by the verbal nature of the contextual problems in mathematics, which is a very relevant finding for assessment practices.

Furthermore, contextual problems did not elicit different solution strategies, contrary to expectations based on research on word problems with young children (Verschaffel et al., 2000) and on algebra problem solving with older students (Koedinger & Nathan, 2004; Koedinger et al., 2008), as well as to expectations based on RME theory (e.g., Van den Heuvel-Panhuizen et al., 2009). However, the absence of effects on strategy choice is more congruent with the findings from Van Putten et al. (2005), who also found no difference in solution strategy choice on multidigit division problems in grade 4.

In the following, we will address the limitations of this study and the implications of its findings, and we will pay special attention to two interesting patterns found in the present study: one regarding gender differences and the other regarding solution strategy use across the four arithmetic operations.

Limitations

The major limitations of the current study lie in the problem set. First and foremost, a rather diverse set of problems and contexts was used in order to obtain a set of problems representative of everyday-life school mathematics. This was done because the ecological validity of the study's findings was deemed more important than a systematic investigation of the effects of different problem characteristics, such as number features, semantic structure, and types of contexts. As a

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

consequence, however, the conclusions may not generalize beyond the types of typical contextual arithmetic problems used in this study, i.e., one-step problems without redundant information or misleading key words. Furthermore, it was not possible to address the effects of specific characteristics of the context, such as linguistic complexity of the problem text (Abedi & Hejri, 2004; Abedi & Lord, 2001) and the effect of an illustration in the problem (Berends & Van Lieshout, 2009).

In addition, due to practical constraints, the number of problems is rather small, in particular when the results per operation are reported. Therefore, these results must be interpreted with caution and should be replicated in future research. However, given the lack of effects of contexts across operations, we believe that it is a rather robust pattern. Furthermore, we found that the (absence of) context effects persisted across different subgroups of students, further supporting the robustness of the pattern.

Furthermore, two types of problem matching were used. The first type was between contextual problems and their corresponding numerical format. However, this correspondence is not univocal, and the possibility that a different choice of matching numerical problems would result in different performance and/or strategy use could not be ruled out. We tentatively argue, however, that for the age group in the current study it may play a minor role given the students' ample experience with switching the order of the numbers and changing the form of a problem.

The second type of problem matching was the construction of parallel problems, in order to prevent testing effects if students had to solve a problem with the exact same numbers twice (once with, and once without a context). Arguably, the parallel problems were probably not exactly parallel in practice because of their unique numerical features. Fortunately, the counterbalancing of parallel problem version and problem format (contextual versus numerical) statistically controls for these possible differences.

Practical implications

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

The results of the current study showed only minor effects of presenting multidigit arithmetic problems in a realistic context at the end of primary school. Because the problems used were taken from the Dutch national assessments at the end of primary school, we tentatively conclude that at least for multidigit addition, subtraction, multiplication, and division problems, the outcomes would have been the same if more or only numerical problems would have been included. This is an important observation, because the mathematics assessments have been criticized for their large demands on language abilities as there are so many verbal problems. In the domain of multidigit arithmetic at the end of primary school, the results of the current study do not seem to support this criticism. Whether these conclusions may be extended to other domains of mathematics (e.g., fractions) and/or other assessments such as TIMSS (grade 4 and grade 8) and PISA (15-year-olds) has to be studied in further research.

Although the presence of a context appeared not to affect arithmetic problem solving at the end of primary school, this does not mean that the shift towards dominance of contextual problems in mathematics tests and mathematics education is without consequences. Research findings in earlier grades (Fuchs et al., 2006, 2008; Hickendorff, in press) did show differences between mathematics problems with and without a context, and also showed that language ability had a larger effect on solving contextual problems. The present results implied that these differences have diminished and disappeared at the end of primary school. However, this does not preclude that students with low verbal abilities had more difficulties to obtain the same performance level on contextual problem solving as on numerical problem solving.

Gender differences

Interestingly, girls outperformed boys on the multidigit arithmetic problems, in contrast to the findings of international (PISA, TIMSS) educational assessments. Dutch national assessments, however, reported the trend that multidigit operations were the only domain of mathematics on which boys did not outperform girls, and in the 2004 assessment, girls even showed a small advantage in this domain (J.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Janssen et al., 2005). A possible explanatory mechanism is differences in strategy use: girls were found to use the more accurate traditional algorithm more often than boys, who in turn were more inclined to answer without written work. This pattern has been reported consistently (Hickendorff et al., 2009, 2010; Hickendorff & Van Putten, 2012) and is also in line with more general research findings concerning gender differences in strategy choice (Carr & Davis, 2001; Carr & Jessup, 1997; Gallagher et al., 2000; Timmermans et al., 2007; Vermeer et al., 2000).

Furthermore, we expected differential effects of the presence of a context for boys and girls, based on findings of Vermeer et al. (2000). This expectation was not confirmed in the current study. A tentative explanatory mechanism may be that grade 6 students (boys and girls) have such well-developed cognitive schemata for solving typical school word problems that they no longer perceive differences in contextual problems and numerical problems. This reasoning is supported by the absence of an effect of the presence of a context for both boys and girls.

Solution strategies for multidigit arithmetic

Although there were no marked effects of the presence of a context on arithmetic problem solving, the present study yields unique empirical data on strategic competence of sixth graders from a reform-based educational environment. That is, strategy use (choice and accuracy) was studied across the four basic operations with multidigit numbers simultaneously, in one common framework. In particular, strategic competence in addition/subtraction on the one hand, and multiplication/division on the other hand, have not previously been studied simultaneously in one study to our knowledge, and several interesting patterns emerged.

First, the dominance of the traditional algorithm decreased from addition (69%) and subtraction (74%) to multiplication (51%) and then again to division (15%). The Dutch national assessments showed a very similar trend in instructional practices: 69%, 72%, 57%, and 17% of the 118 participating grade 6 teachers reported instructing only the traditional algorithm for addition, subtraction,

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

1
2
3 multiplication, and division, respectively (J. Janssen et al., 2005, p. 44). Therefore, it
4 may be that students' choice for the traditional algorithm is determined to a large
5 extent by the teacher's instruction. The traditional algorithm turned out to yield the
6 highest probability of a correct answer on addition, subtraction, and multiplication,
7 but on division it did not differ in accuracy from the non-traditional strategies
8 (similar to findings of Hickendorff et al., 2009). However, because students were
9 free to choose their solution strategy, strategy accuracy data may be biased by
10 selection effects (cf. Siegler & Lemaire, 1997). That is, different students choose
11 different strategies on different items, thereby affecting the accuracy rates. A
12 possible way to assess strategy accuracy without bias would be to implement the
13 choice/no choice method (Siegler & Lemaire, 1997).
14
15
16
17
18
19
20
21
22

23 Second, division stood out compared to the other operations in the frequency
24 of use of non-traditional strategies: 56% of all division trials were solved by
25 repeated subtraction. Again, this is in line with the instructional approach in
26 mathematics textbooks and teacher reports, in which repeated subtraction is the
27 dominant strategy. This may also have been reflected in the strategy accuracy: on
28 each operation except division, non-traditional strategies were less accurate than
29 the traditional algorithm.
30
31
32
33
34
35

36 Third, for each operation, on a substantial percentage of trials (14-17%)
37 students answered without writing down any working. This frequency fell in the
38 range of frequencies reported in previous studies on multidigit multiplication and
39 division (Hickendorff et al., 2009, 2010; Hickendorff & Van Putten, 2012; Van Putten
40 et al., 2005). A previous study showed that it predominantly involved mental
41 computation (i.e., students calculating in their head; Hickendorff et al., 2010).
42 Importantly, it was the least accurate strategy (although the difference with non-
43 traditional strategies was not significant on subtraction), and this is a consistent
44 research finding. Therefore, we plead for more systematic research into this
45 phenomenon, that may yield educational recommendations for instructing students
46 when and when not to use a written strategy.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

References

- Abedi, J., & Hejri, F. (2004). Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Applied Measurement in Education*, 17, 371-392.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14, 219-234.
- Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1-23.
- Adams, R. J., & Wu, M. L. (2000). *PISA 2000 technical report*. Paris: OECD.
- Albert, J. H. (1992). A Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17, 251-269.
- Ambrose, R., Baek, J.-M., & Carpenter, T. P. (2003). Children's invention of multidigit multiplication and division algorithms. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (p. 305-336). Mahwah, NJ: Lawrence Erlbaum Associates.
- Anghileri, J. (1989). An investigation of young children's understanding of multiplication. *Educational Studies in Mathematics*, 20, 367-385.
- Baroody, A. J., & Dowker, A. (Eds.). (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bates, D., & Maechler, M. (2010). *lme4: Linear mixed modeling using S4 classes*. (Computer program and manual). Available from <http://cran.r-project.org/web/packages/lme4/index.html>.
- Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 541-562.
- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *Journal for Research in Mathematics Education*, 24, 294-323.
- Beishuizen, M., Van Putten, C. M., & Van Mulken, F. (1997). Mental arithmetic and

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

strategy use with indirect number problems up to one hundred. *Learning and Instruction*, 7, 87-106.

Berends, I. E., & Van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, 19, 345-353.

Blöte, A., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, 93, 627-638.

Buijs, C. (2008). *Leren vermenigvuldigen met meercijferige getallen* [Learning to multiply with multidigit numbers]. Utrecht, The Netherlands: Freudenthal Institute for Science and Mathematics Education.

Carpenter, T. P., Moser, J. M., & Romberg, T. A. (Eds.) (1982). *Addition and subtraction: A cognitive perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology*, 26, 330-347.

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology*, 89, 318-328.

CITO. (2009). *Terugblik en resultaten 2009. Eindtoets basisonderwijs* [Retrospect and results of the End of Primary School Test 2009]. Arnhem, The Netherlands: Author.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20, 405-438.

De Boeck, P. (2008). Random item models. *Psychometrika*, 73, 533-559.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., et al. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal for Statistical Software*, 39 (12), 1-28.

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A*

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

generalized linear and nonlinear approach. New York, NY: Springer.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77, 460-470.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fischer, G. H. (1987). Linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education. China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, 98, 29-43.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*, 100, 30-47.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, 75, 165-190.

Gravemeijer, K. (1997). Commentary. Solving word problems: A case of modelling? *Learning and Instruction*, 7, 389-397.

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics*, 39, 111-128.

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, 7, 293-307.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Statistical Modelling, 1, 81-102.

Hickendorff, M. (in press). The language factor in elementary mathematics: computational skills and applied problem solving in a multidimensional IRT framework. *Applied Measurement in Education*.

Hickendorff, M., Heiser, W. J., Van Putten, C. M., & Verhelst, N. D. (2009). Solution strategies and achievement in Dutch complex arithmetic: Latent variable modeling of change. *Psychometrika*, 74, 331-350.

Hickendorff, M., & Van Putten, C. M. (2012). Complex multiplication in Dutch educational assessments: What can solution strategies tell us? *Manuscript submitted for publication*.

Hickendorff, M., Van Putten, C. M., Verhelst, N. D., & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology*, 102, 439-452.

Hop, M. (red). (2012). *Balans van het reken-wiskundeonderwijs halverwege de basisschool 5* [Fifth assessment of mathematics education halfway primary school]. Arnhem, The Netherlands: CITO.

Imbo, I., & Vandierendonck, A. (2007). Do multiplication and division strategies rely on executive and phonological working memory resources? *Memory & Cognition*, 35, 1759-1771.

Janssen, J., Van der Schoot, F., & Hemker, B. (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4* [Fourth assessment of mathematics education at the end of primary school]. Arnhem, The Netherlands: CITO.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. D. Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (p. 189-212). New York: Springer.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up. Helping children learn mathematics*. Washington, D.C.: National Academy Press.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13, 129-164.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

- 1
2
3 Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded
4 and abstract representations: Evidence from algebra problem solving. *Cognitive*
5 *Science*, 32, 366-397.
6
7
8
9 Kuss, O., & McLerran, D. (2007). A note on the estimation of the multinomial logistic
10 model with correlated responses. *Computer Methods and Programs in*
11 *Biomedicine*, 87, 262-269.
12
13
14 Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to
15 children's learning of multiplication. *Journal of Experimental Psychology:*
16 *General*, 124, 83-97.
17
18
19 Mabbott, D. J., & Bisanz, J. (2003). Developmental change and individual differences
20 in children's multiplication. *Child Development*, 74, 1091-1107.
21
22
23 Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of
24 multiplication and division. *Journal for Research in Mathematics Education*, 28,
25 309-330.
26
27
28
29 Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics*
30 *report. Findings from IEA's Trends in International Mathematics and Science*
31 *Study at the fourth and eighth grades*. Boston: Boston College, TIMSS & PIRLS
32 International Study Center.
33
34
35
36 National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation*
37 *standards for school mathematics*. Reston, VA: NCTM.
38
39
40 National Council of Teachers of Mathematics. (2000). *Principles and standards for*
41 *school mathematics*. Reston, VA: NCTM.
42
43
44 OECD. (2010). *PISA 2009 results: What students know and can do. Student*
45 *performance in reading, mathematics and science (volume I)*. Paris: OECD.
46
47
48 Prenger, J. (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en*
49 *tekstbegrip in het realistische rekenonderwijs*. [Language counts! A study into the
50 role of linguistic skill and text comprehension in realistic mathematics
51 education]. PhD thesis, University of Groningen, The Netherlands.
52
53
54 R Development Core Team. (2009). *R: A language and environment for statistical*
55 *computing* [Computer software manual]. Vienna, Austria. Available from
56 <http://www.R-project.org>
57
58
59
60

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and Item Response Theory analyses. *Journal of Statistical Software*, 17(5), 1-25.

Robinson, K. M., Arbuthnott, K. D., Rose, D., McCarron, M. C., Globa, C. A., & Phonexay, S. D. (2006). Stability and change in children's division strategies. *Journal of Experimental Child Psychology*, 93, 224-238.

Royal Netherlands Academy of Arts and Sciences. (2009). *Rekenonderwijs op de basisschool. Analyse en sleutels tot verbetering* [Mathematics education in primary school. Analysis and recommendations for improvement]. Amsterdam, The Netherlands: KNAW.

Selter, C. (2001). Addition and subtraction of three-digit numbers: German elementary children's success, methods and strategies. *Educational Studies in Mathematics*, 47, 145-173.

Sherin, B., & Fuson, K. (2005). Multiplication strategies and the appropriation of computational resources. *Journal for Research in Mathematics Education*, 36, 347-395.

Siegler, R. S. (1988a). Individual differences in strategy choices: good students, not-so-good students, and perfectionists. *Child Development*, 59, 833-851.

Siegler, R. S. (1988b). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, 117, 258-275.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, 126, 71-92.

Timmermans, R. E., Van Lieshout, E. C. D. M., & Verhoeven, L. (2007). Gender related effects of contemporary math instruction for low performers on problem solving behavior. *Learning and Instruction*, 17, 42-54.

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM Mathematics Education*, 41, 581-590.

Torbeyns, J., Ghesquière, P., & Verschaffel, L. (2009). Efficiency and flexibility of

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

indirect addition in the domain of multi-digit subtraction. *Learning and Instruction*, 19, 1-12.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2004). Strategy development in children with mathematical disabilities: Insights from the choice/no-choice method and the chronological-age/ability-level-match design. *Journal of Learning Disabilities*, 37, 119-131.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction*, 23, 1-21.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2006). The development of children's adaptive expertise in the number domain 20 to 100. *Cognition and Instruction*, 24, 439-465.

Treffers, A. (1987). Integrated column arithmetic according to progressive schematisation. *Educational Studies in Mathematics*, 18, 125-145.

Treffers, A. (1993). Wiscobas and Freudenthal: Realistic mathematics education. *Educational Studies in Mathematics*, 25, 89-108.

Van den Boer, C. (2003). *Als je begrijpt wat ik bedoel. Een zoektocht naar verklaringen van achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs* [If you get what I mean. A search for explanations of lagging achievement of non-native students in mathematics education]. Utrecht, The Netherlands: CD-β press.

Van den Heuvel-Panhuizen, M., Buys, K., & Treffers, A. (Eds.). (2001). *Kinderen leren rekenen. Tussendoelen annex leerlijnen: Hele getallen bovenbouw basisschool*. [Children learn mathematics]. Groningen, the Netherlands: Wolters-Noordhoff.

Van den Heuvel-Panhuizen, M., Robitzsch, A., Treffers, A., & Köller, O. (2009). Large-scale assessments of change in student achievement: Dutch primary school students' results on written division in 1997 and 2004 as an example. *Psychometrika*, 74, 351-365.

Van Putten, C. M., Van den Brom-Snijders, P. A., & Beishuizen, M. (2005). Progressive mathematization of long division strategies in Dutch primary schools. *Journal for Research in Mathematics Education*, 36, 44-73.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Verhelst, N. D., & Glas, C. A. W. (1995). The one-parameter logistic model. In G. H. Fischer & I. Molenaar (Eds.), *Rasch models. Foundations, recent developments and applications* (p. 215-237). New York: Springer.

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem solving behavior. *Journal of Educational Psychology*, 92, 308-315.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets and Zeitlinger.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*. Greenwich, CT: Information Age Publishing.

Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, 77, 829-848.

Wu, M., & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal*, 18, 93-113.

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

Appendix A

The 8 problem pairs in test form A (texts translated from Dutch).

Note. In the Netherlands, a comma is used as decimal separator.

addition problems

A class of students sells postcards and stamps for charity. They sold for € 466,50 on postcards and for € 985 on stamps. How much did they earn?

$$677,50 + 975 =$$

scarf: €

coat: € 298

shirt: €

hat: € 9,95

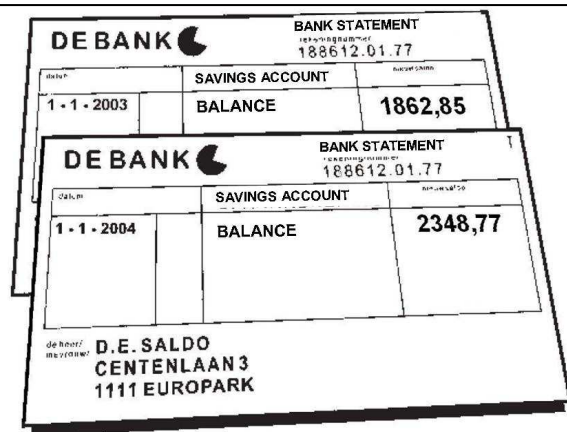
Here, you see the price tags of the clothes that Francien has bought.
How much did she pay in total?

$$19,95 + 198,50 + 129 + 8,80 =$$

subtraction problems

Mrs. De Vries has € 3010 on her bank account. She withdraws € 689 to buy a new bike.
How much is left on her account?

$$4020 - 787 =$$



How much more was there on the account on 1-1-2004 compared to 1-1-2003?

$$3618,88 - 2923,95 =$$

SOLVING MULTIDIGIT MATHEMATICS PROBLEMS

multiplication problems

Charles has to photocopy 36 pages. He needs 27 copies of each page.
How many copies are that in total?

$37 \times 24 =$




Mother buys 17 meters of this curtain material.
How much does she have to pay?

$26 \times 20,1 =$

division problems

Invoice for "The Sunflower"

	number	price per book	total price
history textbooks	32		€ 736

The Sunflower has bought 32 new history textbooks.
How much is the price per book?

$864 : 36 =$



From a 5120 m² piece of land, 16 building parcels are made. Each building parcel will have the same area.
What area will each building parcel have?

$5880 : 14 =$

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Appendix B

Examples of solution strategies; numbers referring to Table 1.

addition	subtraction	multiplication	division	
<p>①</p> $\begin{array}{r} \overset{11}{677,50} \\ + 975 \\ \hline 1652,50 \end{array}$	<p>①</p> $\begin{array}{r} \overset{910}{3010} \\ \overset{210}{\cancel{3010}} \\ - 689 \\ \hline 2321 \end{array}$	<p>①</p> $\begin{array}{r} 36 \\ \times 27 \\ \hline 252 \\ 720+ \\ \hline 972 \end{array}$	<p>①</p> $\begin{array}{r} 32/736 \setminus 23 \\ \underline{64} \\ 96 \\ \underline{96} \\ 0 \end{array}$	
<p>②</p> $\begin{array}{r} 677,50 \\ + 975 \\ \hline 1500 \\ 140 \\ 12 \\ \hline 0,50 \\ \hline 1652,50 \end{array}$	<p>②</p> $\begin{array}{r} 3010 \\ - 689 \\ \hline 3000 \\ - 600 \\ - 70 \\ - 9 \\ \hline 2321 \end{array}$	<p>②</p> $\begin{array}{r} 36 \\ \times 27 \\ \hline 42 \\ 210 \\ 120 \\ \hline 600+ \\ \hline 972 \end{array}$	<p>②</p> $\begin{array}{r} 736 \\ \underline{640-} \quad 20x \\ 96 \\ \underline{96-} \quad 3x \\ 0 \quad 23x \end{array}$	<p>③</p> $\begin{array}{r} 736 \\ \underline{320-} \quad 10x \\ 416 \\ \underline{320-} \quad 10x \\ 96 \\ \underline{64-} \quad 2x \\ 32 \\ \underline{32-} \quad 1x \\ 0 \quad 23x \end{array}$
<p>③</p> $\begin{array}{l} 677,50 + 900 = 1577,50 \\ 1577,50 + 70 = 1647,50 \\ 1647,50 + 5 = 1652,50 \end{array}$	<p>③</p> $\begin{array}{l} 3010 - 600 = 2410 \\ 2410 - 80 = 2330 \\ 2330 - 9 = 2321 \end{array}$	<p>③</p> $\begin{array}{l} 36 \times 20 = 720 \\ 36 \times 7 = 252 \\ 720 + 252 = 972 \end{array}$		
<p>④</p> $\begin{array}{l} 600 + 900 = 1500 \\ 70 + 70 = 140 \\ 7,50 + 5 = 12,50 \\ 1500 + 140 + 12,50 = \\ 1652,50 \end{array}$	<p>④</p> $\begin{array}{l} 3000 - 600 = 2400 \\ 2400 - 80 = 2320 \\ 2320 + 10 - 9 = 2321 \end{array}$	<p>④</p> $\begin{array}{l} 30 \times 20 = 600 \\ 30 \times 7 = 210 \\ 6 \times 20 = 120 \\ 6 \times 7 = \underline{42+} \\ 972 \end{array}$	<p>④</p> $\begin{array}{l} 10 \times 32 = 320 \\ 20 \times 32 = 640 \\ 3 \times 32 = 96 \\ 23 \times 32 = 736 \end{array}$	<p>⑤</p> $\begin{array}{l} 2 \times 32 = 64 \\ 4 \times 32 = 128 \\ 8 \times 32 = 256 \\ 16 \times 32 = 512 \\ 20 \times 32 = 640 \\ 22 \times 32 = 704 \\ 23 \times 32 = 736 \end{array}$
	<p>⑤</p> $\begin{array}{l} 689 + 11 = 700 \\ 700 + 300 = 1000 \\ 1000 + 2010 = 3010 \\ 11 + 300 + 2010 = 2321 \end{array}$	<p>⑤</p> $\begin{array}{l} 2 \times 27 = 54 \\ 27 \quad 4 \times 27 = 108 \\ 27 \quad 8 \times 27 = 216 \\ \dots 16 \times 27 = 432 \\ \underline{27+} 32 \times 27 = 864 \\ 972 \quad 864 + 108 = 972 \end{array}$		

Footnotes

¹ We defined *word problems* as problems containing only text, while the category *contextual* problems encompasses word problems, but also contains problems that include an illustration that may hold essential information for problem solving. In this study, we focus on the more general contextual problems.

² We cross-checked this latent correlation estimate by using other estimation methods and software. In a first approach, item parameter were estimated for each dimension separately using conditional maximum likelihood (Verhelst & Glas, 1995) and the latent correlation between the dimensions was estimated in a separate step. Second, we used a Bayesian framework: the MIRT models were formulated as normal-ogive instead of logistic models, and parameters were estimated using an MCMC-procedure (see also Albert, 1992 for unidimensional IRT models and Béguin & Glas, 2001 for MIRT models), that was programmed into R (R Development Core Team, 2009)). Finally, we used the NLMIXED-procedure from SAS (see De Boeck & Wilson, 2004 and Hickendorff, in press) to fit the 2PL extension of the MIRT model, allowing for the nonzero discrimination parameters to be different from each other. In all three alternative programs, the latent correlation was estimated at 1.000, so we conclude that it is a robust result.

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Table 1

Solution strategies categories.

	addition	subtraction	multiplication	division
1	traditional algorithm	traditional algorithm	traditional algorithm	traditional algorithm
2	RME approach	RME approach	RME approach	repeated subtraction (HL)
3	partitioning 1 operand	partitioning 1 operand	partitioning 1 operand	repeated subtraction (LL)
4	partitioning ≥ 2 operands	partitioning both operands	partitioning both operands	repeated addition (HL)
5	-	indirect addition	repeated addition	repeated addition (LL)
6	other written strategy	other written strategy	other written strategy	other written strategy
7	no written working	no written working	no written working	no written working
8	wrong procedure	wrong procedure	wrong procedure	wrong procedure
9	unclear strategy	unclear strategy	unclear strategy	unclear strategy
10	skipped problem	skipped problem	skipped problem	skipped problem

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Table 2
Distribution of solution strategy categories of numerical (num) problems and contextual (context) problems, in proportions per operation. Strategy categories refer to Table 1.

strategy	addition		subtraction		multiplication		division	
	num	context	num	context	num	context	num	context
1 (traditional)	.68	.70	.74	.73	.52	.51	.15	.14
2 *	.09	.09	.01	.01	.03	.02	.53	.50
3 *	.01	.01	.01	.00	.18	.17	.05	.04
4 *	.04	.03	.01	.01	.07	.07	.03	.03
5 *	n.a.	n.a.	.02	.04	.01	.02	.01	.01
6 *	.02	.01	.01	.01	.01	.01	.02	.03
7 (no written)	.14	.14	.17	.17	.16	.16	.14	.17
8 (wrong)	.00	.00	.00	.00	.00	.01	.01	.02
9 (unclear)	.01	.01	.01	.01	.01	.01	.02	.02
10 (skipped)	.01	.00	.01	.01	.01	.02	.04	.04
N observations	1300	1300	1300	1300	1300	1300	1300	1300

* operation-specific strategy categories, see Table 1.

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Table 3

Strategy choice distribution (in proportions), by gender and language achievement level.

strategy	gender		language achievement		
	boy	girl	low	medium	high
traditional	.45	.59	.47	.55	.57
non-traditional	.29	.29	.29	.29	.29
no written working	.22	.09	.19	.14	.12
other	.04	.03	.06	.02	.02
<i>N</i> observations	4768	5216	3776	3712	2608

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Figure Captions

Figure 1. Design of experimental task forms. A = Addition, S = Subtraction, M = Multiplication, and D = Division. Problem indices 1 (small numbers) and 2 (large numbers) denote the specific pair within each operation, indices *a* and *b* denote the two parallel versions within each problem pair. Problems in unshaded cells represent numerical problems, problems in cells shaded grey are the contextual problems.

Figure 2. Graphical representation of between-item two-dimensional IRT model.

Figure 3. Strategy choice proportion of recoded solution strategies on numerical (num) and contextual (context) problems, per operation.

Figure 4. Estimated mean accuracy of the three strategies, by operation.

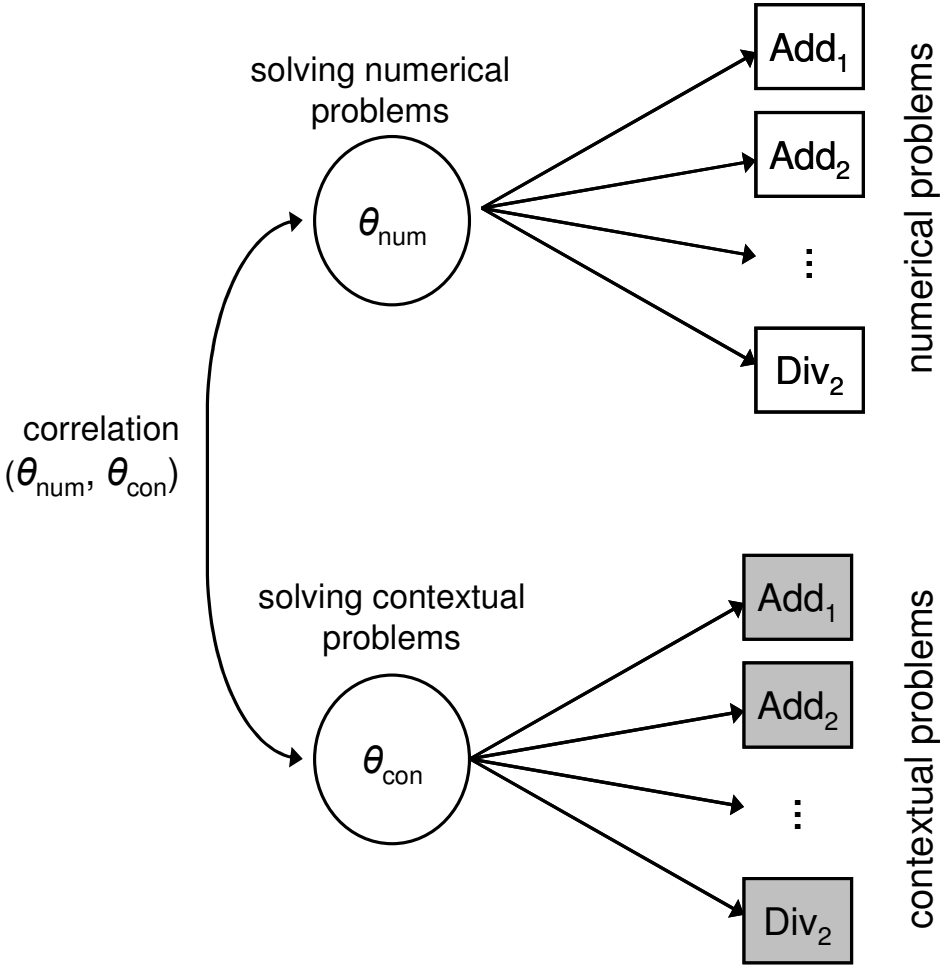
THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

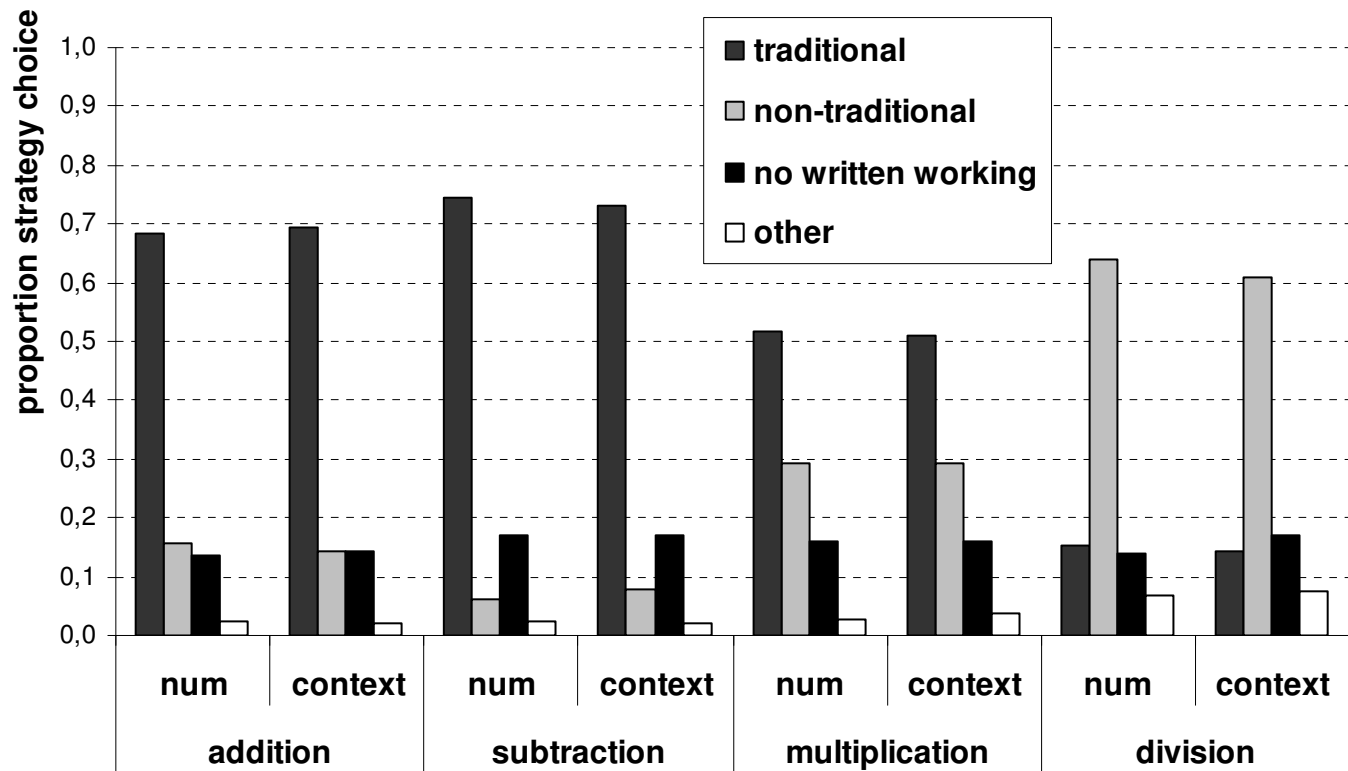
Figure 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
form A	A _{1b}	S _{2a}	A _{2a}	S _{1b}	M _{1b}	D _{2a}	M _{2a}	D _{1b}	A _{1a}	S _{2b}	A _{2b}	S _{1a}	M _{1a}	D _{2b}	M _{2a}	D _{1b}
form B	A _{1b}	S _{2a}	A _{2a}	S _{1b}	M _{1b}	D _{2a}	M _{2a}	D _{1b}	A _{1a}	S _{2b}	A _{2b}	S _{1a}	M _{1a}	D _{2b}	M _{2a}	D _{1b}

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Figure 2



*THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING***Figure 3**

THE EFFECTS OF CONTEXTS ON MATHEMATICS PROBLEM SOLVING

Figure 4

